自引研究综述:科学评价与科学交流中的质疑、求证与创新*

■ 温芳芳
河南科技大学管理学院 洛阳 471023

摘要:[**目的/意义**]自引本是科学交流的一种普遍现象,但在科学评价问题上却陷入了长久的争议。自引研究综述有助于增进学者们对自引的认识和了解,理清对自引的偏见与误解,启发更多人对自引做持续的关注与思索。[**方法/过程**]通过国内外相关文献的系统调研分析,客观描述自引研究的发展历程与现状,梳理其学术脉络和演化轨迹,归纳主要成果和思想,指出当前研究存在的问题,预测未来研究重点与方向。[**结果/结论**]自引研究经历了长期的质疑和反复的求证,至今并无定论,分歧源于研究视角的差异。自引研究亟需新的突破,重心将从单纯的计量与统计分析转向数据背后规律和机理的深度挖掘与解析,自引在考察学术传承和知识扩散方面的功能也将被进一步发掘。

关键词: 自引 科学交流 科学评价 引文指标
分类号: G250
DOI:10.13266/j.issn.0252-3116.2019.21.012

1 引言

自引本是科学交流中的一个自然现象,却在科学评价中饱受质疑和争议。E. Garfield 说"自引本无所谓好坏"[1],关键在于如何看待和利用它。科学借助引文体系呈现出清晰的学术轨迹与脉络,在体现科学研究的连贯性与继承性时,自引与他引并无差别[2]。但随着引文指标越来越多地应用于科学评价活动中,期刊遴选、成果发表、基金评选、职称晋升、科研奖励等严重依赖于引文指标。引文与利益重重捆绑,当引文指标积蓄了如此大的能量,它们是否还能客观地反映科学的真实面貌,是否会在人为操纵之下偏离原有的轨道,学界对此表示出忧虑和怀疑。而相较于他引,自引更容易受到人为操控,所以首当其冲地受到学者们的声讨和质疑。几十年间国内外涌现出一大批关于自引的文章,有人为自引"正名",也有人为其添加新"罪证"。这场旷日持久的争论,至今仍无定论。

众多文献计量学家凭借对计量和评价的敏锐性与严谨性,大胆设想小心求证,全方位多角度地计算和检验自引在科学交流中的普遍性与特殊性、自引对科学评价的影响方式及程度,也有学者另辟蹊径通过自引关系考察知识交流与扩散的规律,以及学术传承的脉络与变迁。本文以"self-citation""self-cited""self-citing""自引""自被引""自引证"作为检索词,分别在 Web of Knowledge、Google Scholar 和中国知网(CNKI)数据库中进行主题检索获得自 20 世纪 70 年代至今的相关文献,再根据篇后所附的参考文献进行扩展检索,共获得外文文献 140 余篇、中文文献 60 余篇。笔者对这些文献进行广泛阅读和系统梳理,归纳和总结主要观点与发现,并征询一些专家学者的意见与建议。这些成果如实记录了学者们几十年如一日对自引问题执着的探索,因关注而质疑,因质疑而求证,因求证而明朗,因明朗而期望,因期望而创新。成果中包含了大量有价值的思路、观点、方法、数据与发现,撷其精华,穿缀成研究综述,梳理自引研究的发展历程与现状,预测未来的趋势与方向。

2 自引:科学交流的必然结果

2.1 自引的普遍性

自引通常指施引和被引文献拥有相同作者的情况

* 本文系国家社会科学基金项目"基于多维信息计量分析的学术影响力综合评价研究"(项目编号:15CTQ023)和河南省哲学社会科学规划项目"基于专利家族计量的企业技术布局及扩张路径研究"(项目编号:2018CZH012)研究成果之一。
作者简介:温芳芳(ORCID:0000-0002-4679-2520),副教授,博士,硕士生导师,E-mail:wenfangfang2013@163.com。
收稿日期:2019-02-14 修回日期:2019-05-21 本文起止页码:117-127 本文责任编辑:王传清

(作者自引),该定义也可引申至其他引文关联,例如,施引和被引文献来自相同期刊(期刊自引)、相同学科(学科自引)、相同语言(语种自引)、相同机构(机构自引)、相同国家(国家自引)。其中,作者自引和期刊自引最受关注。自引是引文系统的有机组成部分,与其他的引文现象一样,自引可以借助数学工具和模型进行分析[3]。论文发表后,其包含的参考文献是固定的,而获得的被引频次却在动态的累积中,所以,S. M. Lawani 建议从共时和历时两个角度考察,将自引分为共时自引(synchronous)和历时自引(diachronous)两种,分别采用自引证率(self-citing rate)和自被引率(self-cited rate)两个指标测度[4]。作者自引证率是指某作者发表的文章中自引参考文献占参考文献总量的比例,自被引率是指某作者的自引频次在其获得的总被引频次中的比例。ISI 将期刊的自引证率定义为某期刊包含的全部参考文献中来自本期刊的参考文献所占比重,将自被引率定义为某期刊获得的总被引频次中来自本期刊的引用次数所占的比重。例如,某期刊自我引用 a 次,引用其他期刊 b 次,被其他期刊引用 c 次,那么该期刊的自引证率是 a/(a+b),自被引率是 a/(a+c)。

早期研究多是计算自引率指标以考察自引的普遍性。E. Garfield 指出仅以第一作者计算,自引在全部引文中所占比例约为 10%,若全部合著者都计算在内该比例会更大[5]。M. H. MacRoberts 认为自引在全部引文中所占比例约为 10% –30%[6]。M. L. Wallace 等的统计结果表明,自引的比例约为 10% –36%,学科之间存在着明显差异[7]。H. Snyder 和 S. Bonzi 的统计结果表明全部引文中约 9% 为自引,物理学、社会科学和人文类学科的自引率分别为 15%、6% 和 3%[8]。R. Tagliacozzo 发现植物生理学和神经生物学的自引率分别为 16.6% 和 17.5%[9]。M. Y. Tsay 以半导体领域高发文量期刊为样本,计算出自引证率和自被引率的平均值分别为 9.59% 和 15.03%[10]。M. Leblond 以生态学领域期刊为例计算自引率为 12.8%[11]。W. B. Lievers 和 A. K. Pilkey 统计了经济学、计算科学、药学与生物学三类学科的 9 种期刊,自引文章的比例为 62.4% –95.1%,平均值为 76.2%;不同期刊的自引率在 4.8% –13.4% 之间,平均值为 9.9%;三类学科的平均自引率分别为 5.9%、11.0% 和 12.4%[12]。

以上数据表明自引率因学科而异、因期刊而异、因统计时间、方法和样本量而异,自引在整个引文系统中常见的比例约为 10% –40%[13]。并且统计中发现,除了学科差异以外,自引率还受到合著、发文量、年龄、引文窗口、语言、出版类型等多种因素的影响。S. Mishra 等归纳和比较了各种因素,认为自引主要受制于文献及其作者和期刊的可见度和可获得性,文化或性别方面的影响是次要的[14]。多方统计结果证实,自引确实是一种普遍的现象,是科学交流的重要组成部分。这也恰恰成为争论的焦点所在:一种观点认为自引率如此之高,必然会对引文评价指标产生显著影响,导致评价结果扭曲或失真,所以在科学评价过程中应该剔除自引;另一种观点则认为自引的比重如此之大,剔除自引显然会破坏引文体系的完整性,使评价结果呈现出另一个极端的偏差和误导。

2.2 自引的动机

A. N. Larcombe 和 S. C. Voss 认为自引是合理和必要的,作者借鉴自己前期成果中的方法、观点和结论来支撑当前研究,尤其一些比较特殊或相对独立的专业领域,难以从外部获得过多的参考文献,必然要大量自引,如果出于上述原因自引并无不妥,且与他引动机没有明显差别[15]。P. Pichappan 和 S. Sarasvady 则认为自引的动机不同于他引,大致包括:延续之前的研究工作、提高早期成果的可见度、提示后续的研究、对低被引或零被引的早期成果进行自我宣传、为自己之前不够完善或者存在争议的观点补充新论据、给审稿人和编辑留下好印象、让读者对自己的研究更有信心和期望、人为放大自己的被引频次等[16]。

J. Hartley 列举了几种自我强化动机:自我标榜——有意告诉审稿人和读者自己之前在高水平期刊上发表过论文;自我宣传——引用自己之前出版的著作意在提升销量;自我推介——引用自己之前发表的文章以提升其可见度,继而获得其他学者的关注和引用[17]。诸如此类的自引,无论施引与被引文献之间是否存在相似性和相关性,其引用动机的合理性都值得怀疑。至于某些期刊和作者为了提升自己的被引频次而刻意自引,以便在科学评价结果中有更好的表现,进而获得与自己真实能力和水平并不相符的地位和荣誉,凡出于这样功利目的的自引就完全可以判定为不当自引[18]。

K. Hyland 和 F. Jiang 把自引动机归为两大类:一是科学研究继承性与连贯性的客观需要;二是作者或期刊刻意提升引文表现的主观需求[19]。前者是合理自引,而后者则被视为不当自引,对此学者们并无异议,但问题在于自引的真实动机难以识别[20]。很多时候,合理自引与不当自引并非泾渭分明,不当自引行为

具有一定的隐蔽性,除作者本人以外,他人很难准确地区分和判定。国内学者朱大明总结了合理自引与不当自引的判断标准,认为自引合理与否关键在于学术诚信[21]。

3 自引:科学评价的干扰因素

3.1 自引与期刊影响力评价

3.1.1 提升期刊影响力的策略性手段 E. Garfield 于1955年提出影响因子指标(Journal Impact Factor, JIF),原为帮助图书馆员和科学家提升采购和检索效率,1972年首版期刊影响因子排行榜问世,随后JIF被广泛应用于科学评价。尽管学界对此褒贬不一,但是正如C. J. A. Bradshaw 和 B. W. Brook 所言,无论你喜欢与否,利用JIF评价期刊及相关科研绩效的做法都是客观且普遍的存在;无论它正确与否,作者投稿、期刊声望、出版商业绩都与JIF息息相关[22]。JIF计算方法中的一些天然漏洞,导致其分子和分母都很容易被操纵,尤其当JIF越来越紧密地与现实利益捆绑时,人们对其科学性和有效性表示出极大的怀疑[23]。甚至科学家的研究模式与期刊的出版策略也发生了相应的改变,为了更高的学术声望,也可能由于出版商的经济压力,部分期刊不惜采取一些策略性的方案来提升JIF:一是人为增加自引,大量引用该期刊在最近两年发表的文章——作者为取悦编辑自引或编辑授意作者自引,无论主动或者被动,理论上都有助于提升JIF;二是部分期刊结成"互引俱乐部",大量引用彼此的文献,这种行为比自引更隐蔽,即便剔除自引对其也无可奈何;三是大量刊登 editorial、letter 等类型的文章,因为影响因子分母只计算 article 和 review 的数量,而分子却将所有类型文章获得的被引频次都计入其中;四是有倾向性地筛选稿件,优先录用那些在编辑看来更易于获得引用的知名学者、机构以及国际合著论文。

3.1.2 自引对于影响因子指标的影响 理论上自引能够人为放大被引频次、提升影响因子,现实情况是否如此,学者们进行了大量的检验,并得出了各色的结论:

(1)正向影响。J. M. Campanario 计算了剔除自引前后的JIF,发现期刊排名会发生很大的变化[24]。J. M. Campanario 和 L. González 的研究表明不当自引可导致影响因子膨胀30%-40%,包括享有极高声望的期刊也是如此[25]。J. K. Vanclay 证实策略性自引确实能够提升被引频次,但作者强调这只是补充性的方案而非替代性方案,且只在短时期内有效,科学影响力的根本在于论文质量而非其他[26]。

(2)负向影响。M. Mimouni 以儿科期刊为例,证实JIF与自引率指标显著负相关,通过自引操纵JIF的说法缺乏有效的证据[27]。T. F. Frandsen 以社会科学领域期刊为例,发现当期刊自引率增长时,JIF不仅没有增加反而还有所下降[28]。

(3)无影响。A. F. J. van-Raan 认为自引确实是一个不容忽视的现象,但自引并没有放大JIF[29]。J. M. Campanario 对JIF连续多年上涨的期刊进行研究,发现JIF增长并非由自引引起,自引对JIF贡献率很低,没有证据表明过量自引会导致JIF膨胀[30]。M. H. Huang 和 C. W. Y. Lin 以环境工程领域的期刊为例,发现剔除自引前后JIF和即年指标(Journal Immediacy Index, JII)都没有明显变化,JIF与自引率没有显著的相关性[31]。A. W. K. Yeuan 认为JIF与自引率无关,作者建议结合多种指标对期刊影响力进行评价,而不能单纯依赖某一种指标[32]。

(4)有条件的影响。X. D. Xia 和 Y. W. Wu 以222种中文期刊为例发现期刊自引率与JIF在某些年度显著正相关,有些年度不相关,作者认为适量自引是必要的,但不应该受到刻意的鼓励或要求[33]。R. Giri 发现自引对于低水平期刊的影响较为显著,而对高水平期刊的影响很小[34]。

3.1.3 影响因子是否因自引而被操纵 G. Yu 和 L. Wang 指出自引率理论上会对JIF产生影响,但现实的情况与理论上的推断总有一定的出入,通过自引操纵JIF的做法对有些期刊可行,对另一些期刊不可行[35]。J. M. Campanario 和 A. Molina 的研究结论支持JIF的鲁棒性,认为确实存在一些能够提升JIF的策略性方案,但并不常见,JIF仍然是能够反映期刊质量和影响力的有效指标[36]。A. Andrade 等选择JIF增长显著的期刊进行研究,考察其增长是否由自引导致,结果表明大部分期刊的JIF增长并不依赖自引,只有极个别的期刊自引影响因子的增速明显大于他引影响因子的增速,作者认为通过自引人为操纵JIF的说法缺乏有效的证据[37]。刘雪立证实我国的SCI来源期刊不存在明显的人为操纵现象[38]。郑毅的研究结果表明,虽然人为增加自引可以提高JIF,但是效果有限,且对提升期刊质量无益[39]。上述研究只是从整体上否认了JIF操纵行为的普遍性,现实中不当自引现象仍屡有发生。曾经一项调查表明,约有20%的受调查期刊不同程度地卷入不当引用,面对编辑的强制自引要求时只有不到7%的作者会拒绝[40]。J. Krauss 列举了6种有强制

自引行为的期刊,发现当自引行为受到刻意鼓励或强制要求时,引文统计结果必然会因人为操纵而失真,虽然全部期刊的自引率没有普遍增长,但是这6种期刊的自引率确实在显著增长[41]。T. Yu 通过调查获得了48种强制自引的期刊[42]。JCR 在2007－2014年间以"过高的自引率已经影响到 JIF 有效性,导致其无法反映期刊的真实影响力"为由先后剔除了250种期刊[43]。

3.2 自引与个人影响力评价

3.2.1 作者自引是否能够影响 h 指数　自 h 指数诞生伊始,学者们就对 h 指数面临的被操纵风险进行了持续讨论,并相继提出一系列的改进指标,包括 g 指数、w 指数、π 指数、e 指数等多达几十种。期刊自引争论的焦点在于自引对 JIF 的影响及程度,作者自引关注的则是自引对 h 指数的影响及程度。J. E. Hirsch 承认自引可以增加总被引频次,但是由于 h 指数同时取决于发文量和被引量两方面,他认为单纯改变被引量并不能对 h 指数产生显著影响[44]。显然,J. Hirsch 的辩护不足以让人信服,多位学者对 h 指数及其引申指标进行检验,结果表明自引对于 h 指数有着显著的影响,剔除自引后 h 指数平均下降12%,且自引对于 g 指数的影响程度大于 h 指数[45-47]。R. J. C. Brown 计算了剔除自引前后的 h 指数,发现存在明显差别,作者建议计算 h 指数时剔除自引,并提出了相应的 b 指数[48]。G. A. Vîiu 检验了 h 指数及其18种引申指数的鲁棒性,发现每种指数都面临不同程度被操纵的风险,作者建议多个指数组合使用以便相互检验[49]。而 M. H. Huang 和 C. W. Y. Lin 却发现自引对于 h 指数指标值及排名的影响微乎其微[50]。

3.2.2 作者自引对 h 指数的影响程度　问题的关键不在于作者自引是否会影响 h 指数,而是影响程度如何[51]。C. Bartneck 和 S. Kokkelmans 采用机器学习方法虚构两名作者,一位采用非正常手段蓄意增加自引,另一位采用随机策略,然后模拟其逐年发文和被引情况,结果显示20年后两位作者的 h 指数分别达到19和15,说明策略性自引确实能提升 h 指数;策略性自引方案适合于发文量和被引量较低的作者,他们的 h 指数基数小,对自引更敏感[52]。R. H. Gálvel 剔除自引后再计算 h 指数,发现年长学者 h 指数下降20%,而年轻学者的 h 指数下降40%[53]。S. R. Hutson 发现年长作者相较于年轻作者自引频次和自引率更高,因为年长学者前期积累成果较多,更易于自引[54]。此外,学者们还证实了自引的"自我推介"功效能够增加他引:J. H. Fowler 和 D. W. Aksnes 指出每增加1次自引一年后可以增加1次他引,5年后可以增加3次他引,10年后增加3.65次他引;尽管当前的自引比重仅为11%,但是由于累积效应,10年以后总被引频次中将会有40%的引文间接来自于自引[55]。M. H. Medoff 在考察自引对作者后续成果的影响时发现,当引文窗口较短时,自引对他引及总被引频次的影响更加显著,5年以内自引确实能够提升该作者后续成果的被引频次,但是超过5年以后再无影响[56]。即便统计引文指标时剔除自引,也无法消除自引能够增加他引的事实,况且我们根本无法区分哪些他引是由自引带来的。

3.3 旷日持久的质疑与争论

3.3.1 自引是否会导致引文指标扭曲　关于自引的争论主要集中于自引是否能够反映真实的影响力以及能够在多大程度上导致引文指标扭曲,从而影响到引文指标的有效性[57]。H. L. Fang 对过度自引和零自引期刊进行比较,发现零自引期刊并非都是顶级期刊,*Science*、*Nature* 这样的期刊也存在一定比例的自引[58]。J. P. A. Ioannidis 指出,即便自引能够干扰引文指标,并不意味着自引天生就是负面的,很多时候自引是合理的,甚至可以说是必要的[59]。潘云涛和武夷山认为作者自引是不可回避的,就连诺贝尔奖得主们都有一定的自引,尽管其自引量占总被引次数的比例不高;正常的自引是将科学研究工作脉络交代清楚的必然要求;每个学科都有自引的规律性,违背学科规律刻意增加自引或回避自引都不可取[60]。与此同时,现实中不当自引现象屡有发生,例如,某作者在一篇文章中自引26次,某一组作者的自引率达59%[11]。M. Mussard 和 A. P. James 报道了2017年泰晤士高等教育世界大学排名中的一起典型案例,印度 Veltech 大学在亚洲排名43,其被引频次更是居亚洲之首,一切都得益于一位"高被引"教授 S. Vaidyanathan,据统计该校86%的被引频次来自于该名教授,他的总被引频次为15 485,自被引率却高达96%[61]。异常的自引不仅使得个人影响力虚高,也导致了大学评价结果的扭曲。

3.3.2 科学评价中是否应该剔除自引　有学者提出只有他引才能显示出文献及其所属的期刊和作者在科学交流系统中的影响力,自引对于科学影响力评价并无太大价值[62]。P. Pichappan 认为尽管不能完全否认自引的影响力,但至少表明自引的影响力是有限的[63]。考虑到自引不具有外部影响力,再加上自引有被操纵风险,有人建议在科学评价时剔除自引,仅以他引来衡量期刊和作者的影响力[64]。对此做法,也有学

者表示反对:过高的自引率说明科学成果的可见度低、相对孤立、缺乏交流,但是没有自引又在一定程度上表明当前研究缺少合理的出处[65]。现实中鲜有学者从未引用过自己的成果,适当的自引恰好能够说明其学术方向稳定、研究工作具有连续性。试想一种期刊如果连自己的作者都不引用,又如何能吸引他引,否定自引本就是一个悖论,也不符合科学交流的规律[66]。不当自引只是个别现象,对整体没有太大影响,而且考虑到自引的普遍性,自引在整个引文系统中所占比重较大,盲目剔除可能导致另一种形式的失真,仍然无法真实客观地反应科学影响力[67]。

关于是否剔除自引,学者们持有不同的意见,且这些观点都是建立在实证检验的基础之上:S. Carley 证实无论是宏观层面还是微观层面,自引都会对引文评价结果产生显著影响[68]。M. Dunaiski 以计算机领域的杰出科学家为例,证明了自引对 h 指数、g 指数等多个引文指标都有显著的影响,认为剔除自引有助于提升引文指标的准确性[69]。C. G. Yheulon 以微创手术领域的学者为例,证实作者自引对 h 指数等引文指标的影响很小,在个人科研业绩评价时没有必要剔除自引[70]。W. Glänzel 的研究表明从国家层面分析,自引对引文指标的影响不大,没有必要剔除自引[71]。D. Z. Zhao 的研究表明,大部分情况下自引都是出于合理动机,自引较之他引更具计量价值,科学评价不仅不能剔除自引,还应该赋予其更高的权重,作者将剔除自引的做法比喻为"把孩子和洗澡水一起倒掉"[72]。争执之外还存在着相对保守的第三种声音,认为完全剔除或者坐视不理都不可取,建议在计算引文指标时提供包含自引和剔除自引两种方案,这样才能对科学交流的复杂性有更好的理解[13]。目前,JCR 提供包含自引和不含自引两种 JIF 指标,Scopus 也提供包含自引和不含自引两类 h 指数,表明主流数据库对自引同样持有审慎的态度。

3.3.3 如何识别不当自引　通过投诉、举报或社会调查的形式识别不当引用虽有个案,但不具有广泛的适用性[73]。从文献计量学角度出发监测自引模式的异常变化,为不当自引的识别提供了一种较为可行的思路。E. Garfield 曾提出 20% 为合理自引的标准,自引率超过 20% 的期刊为疑似过度自引[74]。JCR 在识别不当自引期刊时,也将自引率 20% 和两年自引率 50% 作为参照。B. Z. ILiev 倡议设置自引频次的上限,高于上限的视为非正常自引,在科学评价中应予以剔除或者降低权重[75]。显然,此种方案并未考虑学科领域的差别性以及某些期刊的特殊性。C. Humphrey 认为正常的自引率应该保持一定的稳定性,可以通过监测自引率指标的历时变化来识别疑似过度自引,而不能仅凭自引率指标的大小作判定[76]。自引率在某个年度突然增长或者多年持续增长,都有过度自引的嫌疑[77]。金铁成建议比较自被引率与两年自被引率,如若两者相差过大,则可视为异常现象,理由是如果没有人为操纵,两个指标应当大致相等,除非为了抬高 JIF 而有意扩大最近两年的自引量,而导致两年自被引率异常[78]。还有学者建议借助专家经验判断从源头上把关,期刊的编辑和审稿人对本专业的文献比较熟悉,建议由他们在审稿过程中识别和纠正不当自引[79]。J. W. Flatt 建议自引透明化,如实报告期刊和作者的自引状况,使其处于学术共同体的监督之下,如此一来,不当自引行为必当有所顾忌,也避免了不当自引现象的蔓延[80]。

4 自引:知识扩散的特殊模式

4.1 自引源于知识生产的继承性

科学出版物本身就是一个巨大的知识扩散系统,引文在研究者获取新知识的过程中发挥着积极的推动作用[81]。随着引文在科学评价中的广泛应用,一些策略性的方案可能会影响到作者引用行为以及引文指标的有效性,但不当自引并非与科学交流相伴而生,而是科学发展到一定阶段,当科学评价过度依赖引文指标时导致的非正常现象。尽管个别作者和期刊会采用一些策略性手段增加自引,但是并不能就此否认自引在考察知识交流方面的科学性和有效性。R. Costas 认为应辩证地看待自引现象,从科学评价角度考察,自引是导致评价结果扭曲的一个因素,而从科学交流角度考察自引是知识交流的天然组成部分;自引虽有一定的自我推荐和传播功效,但科学影响力的根本在于他引而非自引[82]。引文反映了不同文献包含的科学观念、方法等之间的互动,展示了科学演进过程,自引也是如此[83]。作者自引是科学家研究方向稳定性和研究成果连续性的标志,体现出研究工作的延续与扩展[84]。科学期刊作为连续性出版物,理应保持相对稳定的读者群和作者群,学者们长期关注、阅读和引用某些期刊,并将自己撰写的文章发表于这些期刊,既是因为文章的研究主题与期刊选稿主题高度一致,也是投稿作者对该期刊质量和影响力的认可。无论是期刊自引还是作者自引,在不考虑强制自引等非正常情况时,自引都是建立在知识生产的连续性和继承性基础上的

科学交流行为，与他引一样，自引反映了文献之间的知识关联性和内容相似性。

4.2　自引是知识扩散的特殊形式

自引是过去和现在研究的交流与对话，反映了作者认知和创造的过程，也是知识经由作者本人进行扩散的一种特殊形式。R. H. Gálvez 训练一个主题发现模型，分别测度了自引文献之间和他引文献之间的语义相似度，发现前者明显大于后者，说明作者自引确实是出于知识的关联性而非其他原因[53]。国内学者刘桂琴的研究证实，自引文献的平均余弦相似度和平均相关系数最高，文献之间显著正相关；他引文献的文本相似度和平均相关系数较低，可见自引主要是受“内容驱动”[85]。自引便于读者掌握研究进展及学术动态，通过“溯源追踪”把握研究起源与演化轨迹，是知识扩散的一种有效机制。以往学者们过于关注自引在科学评价方面的表现，总是质疑自引对引文指标的干扰，而忽视了自引在知识扩散中的积极功效。

L. Zhou 认为他引过程中可能存在着对他人成果不甚了解情况下的“仓促引用”或者根本没有阅读原始文献而“转引”，实际上这样的引文质量很低，对于研究知识扩散并无益处，作者对于自己的前期成果非常熟悉，所以自引较之他引准确度更高[86]。此外，在知识扩散的过程中自引的时效性也优于他引[87]。R. Rousseau 比较了自引和他引频次的年度分布曲线，认为自引更为新颖及时，10 年后自引发生的可能性很小[88]。作者在引用自己的成果时不必等待漫长的出版周期，也不必从海量文献中进行搜索、阅读和比较，而是直接从自己最熟悉的知识库中选择（含已经出版和尚未出版的文献），一些文献在正式出版前就已经提前进入引文系统。自引大大缩短了引文时滞，在利用引文关系考察科学交流和知识扩散规律时，自引相较于他引更能反映新兴的知识、观点和方法。W. Y. L. Lin 等以环境工程的文献为例，证实了自引文献相对于他引文献更具时效性，作者更倾向于引用自己近期发表的成果，但自引和他引文献年龄分布的差异性正在逐渐缩小，可能的原因是在线出版、预印本等新兴出版模式的出现，加快了知识传播的速度[89]。

4.3　自引彰显出学术传承的脉络

科学研究是一项连续性的活动，也是一个不断更新完善的过程，通常不会有剧烈的变化[16]。一位学者在学术生涯的不同阶段发表一系列的成果，这些先后发表的成果具有方向上的一致性、内容上的关联性和思路上的继承性，尤其是在相同或相近时期发表的成果，彼此间存在着紧密的关联。自引将这些具有一致性、关联性和继承性的文献串联起来，成为记录学者研究活动足迹的完整链条。所以，自引为我们考察某一学者学术生涯的研究轨迹、某一学科领域的知识结构及其演化路径提供了重要的支撑。当某一时期自引量很少甚至为零时，说明作者又转向了新领域或者建立起新的合作关系；当一个学科领域的自引越来越频繁，他引越来越少时，说明一个新的领域正在诞生[67]。利用自引可以探测出作者职业生涯和研究领域的变迁，能够识别正在出现的新兴研究主题以及作者研究兴趣的变化情况。

自引是引文系统的子系统，自引网络是一类特殊的引文网络，它是由某一作者及其合作者发表的论文相互引用而形成，可以用来刻画作者的研究主题，观察作者研究轨迹的变化，识别关键成果[90]。I. Hellsten 指出通过作者自引网络能够更好地发现新的研究主题[91]。J. Y. Lee 认为作者自引网络能够成功识别作者的核心论文和领先成果[92]。魏瑞斌以美国印第安纳大学信息、计算与工程学院为例，构建了该机构内部的自引网络，采用主路径分析方法进行研究主题的识别及演变历程的梳理，作者指出该方法可以推广至国家或地区等其他层次的自引分析[93]。A. El-Aichouchi 和 P. Gorry 在对肿瘤血管生成理论之父 J. Folkman 教授的“睡美人”文献进行研究时，发现“王子”是一篇自引文献，科学发展和知识累积是一个缓慢的过程，对于起初并不被同行认可的“超前”理论，自引能够引起同行的关注，并为之积累更多的证据以消除大家的疑虑，最终唤醒了沉睡多年的理论，这是自引所具备的“自我唤醒”功能[94]。

5　思考与展望

自引研究已有半个世纪之久，质疑与争论伴随始终，虽就一些关键性问题仍未达成一致意见，但为后续研究奠定了坚实基础。从最初简单的否定、到审慎的质疑、再到细致的检验、直到近年来发掘自引在考察科学交流和知识扩散方面的新功能。研究方法和思路不断创新，研究态度更趋严谨和理性。本文梳理几十年间的研究成果，沿时间轴绘制其主要脉络和演化轨迹，见图1。

5.1　透过数据表象揭示深层次的规律和机制

以往的研究大部分都是从表层的数据当中解释自引指标（如自引频次、自引率等）与引文指标（如 JIF、

温芳芳. 自引研究综述:科学评价与科学交流中的质疑、求证与创新[J]. 图书情报工作,2019,63(21):117–127.

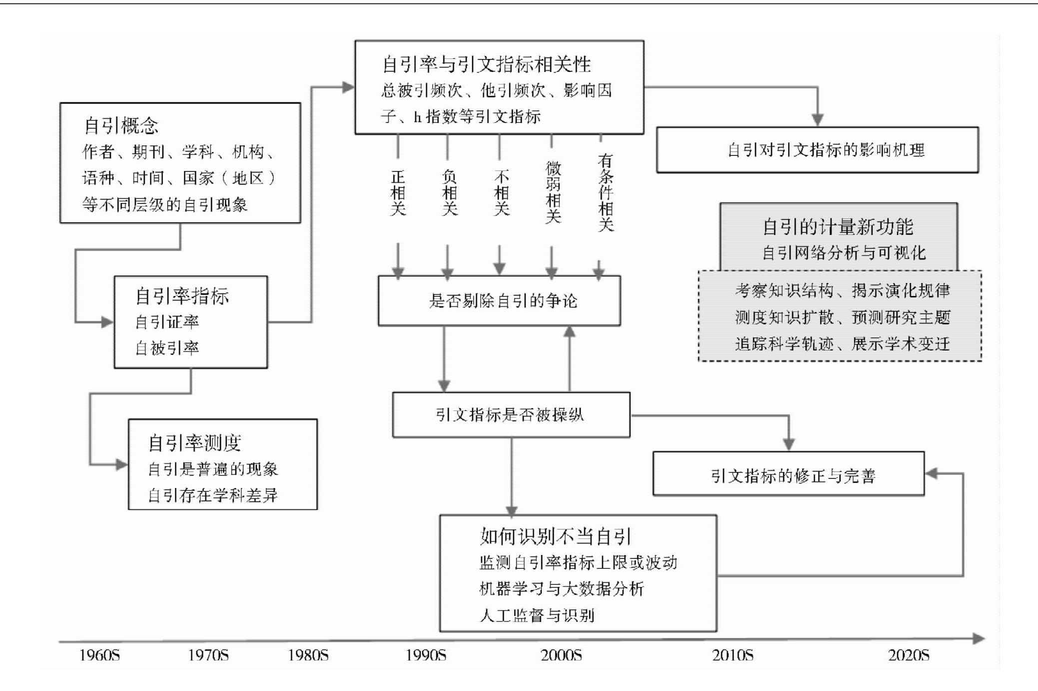


图 1　自引研究的主要演化轨迹

h指数等)呈现出的统计学相关性,凭借统计数据简单地判定是否相关及相关程度,至于其背后的影响机理、作用机制等问题却缺乏深层次的解析。例如,有学者观察到自引能够增加他引频次的事实,然而究竟是因为自引提高了文献及作者的可见度,还是因为自引文献确实具有更高的质量,现象背后的原因仍无法解释。统计数据虽能直接证明自引对于引文指标的影响及程度,但由于样本来源、规模等的局限性,终究无法获得令人信服的结论,甚至在一定程度上加剧了认识上的分歧与混乱,关于"是否应该剔除自引"的争论就是这种分歧和混乱的典型表现。这一过程如同"盲人摸象",似乎每一个人都获得了正确的发现,但实际上只是局部的认识,并未抓住本质,也不能反映整体。

未来研究更应该关注数据背后的东西,从统计数据和量化分析中总结规律和模式,阐释背后的原因。研究的重心不再是自引是否影响引文指标、过度自引是否会导致引文数据膨胀等问题,而是自引如何作用于引文指标、自引与引文指标之间的关联机制和影响机理,只有先理清这些问题才能认清规律、抓住本质,以便提出更有效的应对方案。该领域并不缺少实证研究和量化分析,也不是更大的样本规模、更多学科领域的再检验就能解决问题,真正缺少的是理论上的支撑和依据,迫切需要有学者能够从多年积累的纷繁凌乱的统计数据和观察结果当中,梳理脉络、提炼规律,使大家能够摒弃对自引的偏见和误解,更为客观和理性地认识自引及其可能导致的影响。只有这样,才有可能开启自引研究的新阶段,而不是长久陷入单纯质疑和简单检验的重复循环当中。

5.2　借助自引网络从新的角度考察知识结构

关于自引的争论,根本原因在于大家关注的角度不同。若从科学评价的严谨性角度考察,自引确实存在着人为放大引文指标而导致评价结果失真的风险;若从科学交流的完整性角度考虑,自引是一个不容忽视的科学交流现象。关于自引的是非之争取决于研究者的立场和视角,以往研究大多关注前者而忽视后者。事实上,引文是一种特殊的引文现象,作为科学交流的一种有效机制,肩负着学术传承、知识扩散的重要职责。无论其在科学评价方面的作用是正向还是负向,并不影响自引在考察科学交流、知识扩散等方面所能发挥的积极功效。科学交流的过程中,自引和他引各有侧重,具有互补的功效。科学是一个由作者、文献、期刊等学术主体编织而成的巨大网络,每个主体都能够用来分析与其他主体之间的关联性,自引在考察知识扩散方面所具有的功效值得期待。

在文献计量学领域,引文分析是揭示知识结构的主要手段,包括直接引证、引文耦合、共被引、互引等在内的多种引文分析方法,被广泛地应用于考察知识结构,但是这些研究大多建立在他引的基础之上。自引分析提供了一种新的视角和思路,借助复杂网络分析及可视化技术,有望获得全新的结论和发现。如前文所述,已有研究证实了自引文献的语义相似度大于他引文献、自引较之他引更具准确性和时效性,这些观点

为我们发掘自引的文献计量新功能，拓展其在揭示知识结构、表征学术关系等方面的应用做出了有益的铺垫。未来该研究主题将利用自引脉络追踪作者的学术轨迹，考察作者学术生涯变迁、揭示学科知识结构及其演化规律、预测正在孕育的新兴主题。我们期待着自引分析所能带来的新发现与新结论，当然自引分析的结果也会拿来与他引进行比较，以便进一步明确自引分析在揭示知识结构时的可行性、科学性与特殊性。

5.3 识别不当自引行为以及纠正其影响的方案

尽管我们愿意积极乐观地看待自引，但是也应该持有谨慎的态度。毕竟强制自引、过度自引等情况确实存在，如若坐视不理，可能会导致不当引用行为的蔓延和恶化，最终使自引不能反映真实的科学影响力，更有可能威胁到以引文为媒介的科学交流活动的有序性，而最为恶劣的影响是对学术道德的践踏和对学术氛围的破坏[95]。因此，无论出于保障科学评价准确度、维持学术交流秩序之目的，还是出于制止不当引用、净化学术氛围的考虑，区分合理引用与不当引用、识别不当引用并纠正其在科学评价中的副作用都是十分必要的。以往虽然也有一些研究提出了应对方案，如：设定自引率阈值和自引频次上限、考察自引率年度变化、计算自引文献相似度等。但是这些方案并未获得广泛的认可和应用，其有效性和科学性也有待检验。未来仍需要加强这方面的研究：一方面，利用语义模型、机器学习、复杂网络分析、大数据分析等技术和手段，从海量的自引数据中自动识别出存在或者疑似存在不当自引的期刊、作者和文章；另一方面，在强化学术自律的同时，更需要探讨如何建立不当引用行为的监督约束机制，不仅是不当自引，还包括互引俱乐部、强制引用、馈赠引用等所有不当引用情况，都应该受到学术界共同的抵制和监督。

自引总是处于学术道德的拷问之下，虽没有统一的结论，但终究负面的评论占据上风[96]。自引虽有被操纵的风险，并不意味着自引"本性邪恶"，很多时候只是在不健全的科研评价体系中"误入歧途"。在"不发表就灭亡"(publish or perish)的时代背景下，以引文指标为导向的科研评价制度，建立起引文与学术地位、职业升迁、项目资助、物质奖励之间的紧密关联，在提升科学产出效率的同时，也带来了一些负面效应，不当自引即是其中之一[97]。一些学者甚至认为引文指标根本就不能反映学术成果的质量以及相应学术主体的科学影响力[98]。A. Racz 和 S. Markovic 强烈抨击利用影响因子指标评价期刊、个人、论文、机构的做法，认为所有不当引用行为(包括不当自引)归根结底源于不合理的政策设计，"唯影响因子是从"的评价方法不仅应该被质疑，更应该被彻底废止[99]。再多的争论与求证，即便是找到一种可行的方法能够识别不当引用，也不能从根本上解决问题。从长远计，还是应该改变现行的评价制度，弱化引文指标在科研绩效评价中的作用，积极探索新的评价指标，如此才能从根本上消除滋生不当引用的温床[100]。

致谢：感谢印第安纳大学信息、计算与工程学院 Cassidy R. Sugimoto 博士对本文的指导和帮助。

参考文献：

[1] GARFILED E. Journal citation studies XVII: journal self-citation rates-there's a difference[J]. Essays of an information scientist, 1974,52(2):192-194.

[2] BONZI S, SNYDER H. Motivations for citation: a comparison of self-citation and citations to others[J]. Scientometrics, 1991,21(2):245-254.

[3] GLANZEL W, THIJS B, SCHLEMMER B. A bibliometric approach to the role of author self-citations in scientific communication[J]. Scientometrics,2004,59(1):63-77.

[4] LAWANI S M. On the heterogeneity and classification of author self-citations[J]. Journal of the American Society for Information Science,1982, 33(5): 281-284.

[5] GARFIELD E. Citation indexing-its theory and application in science, technology, and humanities[M]. New York: John Wiley, 1979:245.

[6] MACROBERTS M H, MACROBERTS B R. Problems of citation analysis: a critical review[J]. Journal of the American Society for Information Science, 1989,40(5):342-349.

[7] WALLACE M L, LARIVIERE V, GINGRAS Y. A small world of citations? The influence of collaboration networks on citation practices[J]. PLOS ONE, 2012,7(3):e33339.

[8] SNYDER H, BONZI S. Patterns of self-citation across disciplines (1980-1989)[J]. Journal of information sience,1998,24(6):431-435.

[9] TAGLIACOZZO R. Self-citation in scientific literature[J]. Journal of documentation,1977,33(4):251-265.

[10] TSAY M Y. Journal self-citation study for semiconductor literature: synchronous and diachronous approach[J]. Information processing and management, 2006,42(6):1567-1577.

[11] LEBLOND M. Author self-citations in the field of ecology[J]. Scientometrics,2012,91(3):943-953.

[12] LIEVERS W B, PILKEY A K. Characterizing the frequency of repeated citations: the effects of journal, subject area, and self-citation[J]. Information processing and management, 2012,48(6):1116-1123.

[13] THIJS B, GLANZEL W. The influence of author self-citations on

bibliometric meso-indicators: the case of European universities [J]. Scientometrics, 2006,66(1):71-80.

[14] MISHRA S, FEGLEY B D, DIESNER J, et al. Self-citation is the hallmark of productive authors of any gender[J]. PLOS ONE, 2018,13(9):e0195773.

[15] LARCOMBE A N, VOSS S C. Self-citation: comparison between radiology, European radiology and radiology for 1997-1998[J]. Scientometrics, 2011,87(2):347-356.

[16] PICHAPPAN P, SARASVADY S. The other side of the coin: the intricacies of author self-citations[J]. Scientometrics, 2002,54(2): 285-290.

[17] HARTLEY J. To cite or not to cite: author self-citations and the impact factor[J]. Scientometrics,2012, 92(2):313-317.

[18] OPTHOF T. Inflation of impact factors by journal self-citation in cardiovascular science[J]. Netherlands heart journal, 2013,21(4): 163-165.

[19] HYLAND K, JIANG F. Changing patterns of self-citation: cumulative inquiry or self-promotion? [J]. Text & talk, 2018,38(3): 365-387.

[20] ZAGGL M A. Manipulation of explicit reputation in innovation and knowledge exchange communities: the example of referencing in science[J]. Research policy, 2017,46(5):970-983.

[21] 朱大明. 参考文献"合理自引"与"不当自引"的区分标准[J]. 编辑学报,2004,16(1):76.

[22] BRADSHAW C J A, BROOK B W. How to rank journals[J]. PLOS ONE, 2016,11(3):e0149582.

[23] MARTIN B R. Editors' JIF-boosting stratagems-which are appropriate and which not? [J]. Research policy,2016,45(1):1-7.

[24] CAMPANARIO J M. The journal citation reports (SCI edition) with and without journal self-citation[J]. Scientometrics, 2018,27(2):1699-2407.

[25] CAMPANARIO J M, GONZALEZ L. Journal self-citations that contribute to the impact factor: documents labeled "editorial material" in journals covered by the Science Citation Index[J]. Scientometrics,2006,69(2):365-386.

[26] VANCLAY J K. Factors affecting citation rates in environmental science[J]. Journal of informetrics,2013,7(2):265-271.

[27] MIMOUNI M, RATMANSKY M, SACHER Y, et al. Self-citation rate and impact factor in pediatrics[J]. Scientometrics,2016,108(3): 1455-1460.

[28] FRANDSEN T F. Journal self-citations-analyzing the JIF mechanism[J]. Journal of informetrics,2007,1(1):47-58.

[29] VAN-RAAN A F J. The influence of international collaboration on the impact of research results. Some simple mathematical considering the role of self-citations[J]. Scientometrics, 1998,42(3):423-428.

[30] CAMPANARIO J M. Journal that rise from the fourth quartile to the first quartile in six years or less: mechanism of change and the role of journal self-citations[J]. Publications,2018,47(6):1-15.

[31] HUANG M H, LIN C W Y. The influence of journal self-citations on journal impact factor and immediacy index[J]. Online information review, 2012,36(5):639-654.

[32] YEUAN A W K. Higher impact factor of Neuroimaging journals is associated with larger number of articles published and smaller percentage of uncited articles[J]. Frontiers in human neuroscience, 2019,12(1):1-4.

[33] XIA X D, WU Y W. Journal self-citation analysis of some Chinese sci-tech periodicals[J]. Serials review, 2011,37(3):171-173.

[34] GIRI R. Influence of selected factors in journals' citations[J]. Aslib journal of information management, 2019,71(1):90-104.

[35] YU G, WANG L. The self-cited rate of scientific journals and the manipulation of their impact factors[J]. Scientometrics, 2007,73(3):321-330.

[36] CAMPANARIO J M, MOLINA A. Surviving bad times: the role of citations, self-citations and numbers of citable items in recovery of the journal impact factor after at least four years of continuous decreases[J]. Scientometrics, 2009,81(3):859-864.

[37] ANDRADE A, GONZALEZ J R, CAMPANARIO J M. Journals that increase their impact factor at least fourfold in a few years: the role of journal self-citations[J]. Scientometrics, 2009,80(2):515-528.

[38] 刘雪立,魏雅慧,盛丽娜,等. 科技期刊总被引频次和影响因子构成中的自引率比较——兼谈影响因子的人为操纵倾向[J]. 编辑学报,2017,29(6):602-606.

[39] 郑毅. 科技期刊人为提高自引率对影响因子的定量研究[J]. 江汉大学学报(自然科学版),2012,40(4):52-53.

[40] WILHITE A W, FONG E A. Coercive citation in academic publishing[J]. Science, 2012,335(6068):542-543.

[41] KRAUSS J. Journal self-citation rates in ecological sciences[J]. Scientometrics, 2007,73(1):79-89.

[42] YU T, YU G, WANG M Y. Classification method for detecting coercive self-citation in journals[J]. Journal of informatics, 2014,8(1):123-135.

[43] LIN W Y C. The performance of Asian S&T journals in international citation indicators[J]. Learned publishing, 2017,30(3):193-204.

[44] HIRSCH J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005,102(46): 16569-16572.

[45] KELLY C D, JENNIONS M D. The h-index and career assessment by numbers[J]. Trends in ecology and evolution, 2006,21(4): 167-170.

[46] SCHREIBER M. Self-citation corrections for the Hirsch index[J]. Europhysics letters, 2007,78(3):1-6.

[47] SCHREIBER M. The influence of self-citation corrections on Egghe's g-index[J]. Scientometrics,2008,76(1):187-200.

[48] BROWN R J C. A simple method for excluding self-citation from

the h-index: the b-index[J]. Online information review, 2009,33(6):1129-1136.
[49] VIIU G A. A theoretical evaluation of Hirsch-type bibliometric indicators confronted with extreme self-citation[J]. Journal of Informetrics,2016,10(2):552-566.
[50] HUANG M H, LIN C W Y. Probing the effect of author self-citations on h index: a case study of environmental engineering[J]. Journal of information science,2011,37(5):453-461.
[51] ZHIVOTOVSKY L A, KRUTOVSKY K V. Self-citation can inflate h-index[J]. Scientometrics,2008,77(2):373-375.
[52] BARTNECK C, KOKKELMANS S. Detecting h-index manipulation through self-citation analysis[J]. Scientometrics, 2011, 87(1): 85-98.
[53] GALVEZ R H. Assessing author self-citation as a mechanism of relevant knowledge diffusion[J]. Scientometrics, 2017,111(3):1801-1812.
[54] HUSTON S R. Self-citations in Archaeology: age, gender, prestige, and the self[J]. Journal of archaeological method and theory, 2006,13(1):1-18.
[55] FOWLER J H, AKSNES D W. Does self-citation pay? [J]. Scientometrics,2007,72(2):427-437.
[56] MEDOFF M H. The efficiency of self-citations in economics[J]. Scientometrics, 2006,69(1):69-84.
[57] GLANZEL W, THIJS B. The influence of author self-citations on bibliometric macro indicators[J]. Scientometrics, 2004,59(3):281-310.
[58] FANG H L. Self-citation rates of scientific and technical journal in SCI from China, Japan, India, and Korea[J]. Learned publishing, 2013,26(1):45-49.
[59] IOANNIDIS J P A. A generalized view of self-citation: direct, co-author, collaborative, and coercive induced self-citation[J]. Journal of psychosomatic research. 2015,78(1):7-11.
[60] 潘云涛,武夷山. 自引、他引:说不尽的故事[J]. 科技导报,2007(24):85.
[61] MUSSARD M, JAMES A P. Boosting the ranking of a university self-citations[J]. Current science,2017,113(10): 1827.
[62] 蒋鸿标. "自引文献"不宜提倡[J]. 图书与情报,2004(6):94-95.
[63] PICHAPPAN P. A dual refinement of journal self-citation measures[J]. Scientometrics,1995,33(1):13-21.
[64] FERRARA E, ROMERO A E. Scientific impact evaluation and the effect of self-citations: mitigating the bias by discounting the h-index[J]. Journal of the American Society for Information Science and Technology, 2013,64(11):2332-2339.
[65] TAL D, GORDON A. Publication attributes of leadership: what do they mean? [J]. Scientometrics,2017,112(3):1391-1402.
[66] 朱大明. 他引或自引:关键是学术诚信[J]. 科技导报,2008(16):99.
[67] AKSNES D W. A macro study of self-citation[J]. Scientometrics, 2003, 56(2):235-246.
[68] CARLEY S, PORTER A L, YOUTIE J. Toward a more precise definition of self-citation[J]. Scientometrics, 2013,94(2):777-780.
[69] DUNAISKI M, GELDENHUYS J, VISSER W. Author ranking evaluation at scale[J]. Journal of informetrics, 2018,12(3):679-702.
[70] YHEULON C G, BALLA F M, PATEL A D, et al. Publication patterns and the impact of self-citation among minimally invasive surgery fellowships[J]. American journal of surgery, 2019,217(2):346-349.
[71] GLANZEL W, THIJS B. Does co-authorship inflate the share of self-citations? [J]. Scientometrics,2004,61(3):395-404.
[72] ZHAO D Z, STROTMANN A, CAPPELLO A. In-text function of author self-citations: Implications for research evaluation practice [J]. Journal of the Association for Information Science and Technology,2018,69(7):949-952.
[73] YU T, YU G, SONG Y, et al. Toward the more effective identification of journals with anomalous self-citation[J]. Malaysian journal of library & information science 2018,23(2): 25-46.
[74] GARFIELD E, WELLJAMS D. Citation data: their use as quantitative indicators for science and technology evaluation and policy-making[J]. Science and public policy,1992,19 (5):321-327.
[75] I LIEV B Z. Measuring the evaluation and impact of scientific works and their authors[J]. Journal of geometry and symmetry in physics, 2015,40(2):69-119.
[76] HUMPHREY C, KISELEVA O, SCHLEICHER T. A time-series analysis of the scale of coercive journal self-citation and its effect on impact factors and journal rankings[J]. European accounting review, 2019,28(2):335-369.
[77] CAMPANARIO J M. Large increases and decreases in journal impact factors in only one year: the effect of journal self-citations[J]. Journal of the American Society for Information Science and Technology, 2011,62(2):230-235.
[78] 金铁成. 采用自被引率与2年自被引率检测学术期刊过度自引的比较与分析[J]. 中国科技期刊研究,2016,27(9): 949-952.
[79] 金铁成. 学术期刊自引率使用乱象及其应对策略[J]. 科技与出版,2016(11):96-98.
[80] FLATT J W, BLASIMME A, VAYENA E. Improving the measurement of scientific success by reporting a self-citation index[J]. Publications, 2017,5(3):20-25.
[81] WOJICK D E, WARNICK W L, CARROLL B C, et al. The digital road to scientific knowledge diffusion[J]. D-Lib magazine, 2006,12(6):1082-9873.
[82] COSTAS R, VAN-LEEUWEN T N, BORDONS M. Self-citations at the meso and individual levels: effects of different calculation methods[J]. Scientomerics, 2010,82(3):517-537.
[83] LIU Y X, ROUSSEAU R. Interestingness and the essence of cita-

tion[J]. Journal of documentation, 2013,69(4):580－589.

[84] MUBIN O, ARSALAN M, MAHMUD A. Tracking the follow-up of work in progress papers[J]. Scientometrics, 2018,114(3):1159－1174.

[85] 刘桂琴. 基于作者自引的知识扩散分析[J]. 情报杂志,2018,37(7):146－149.

[86] ZHOU L, AMADI U, ZHANG D. Self-citation biased? An investigation via the lens of citation polarity, density, and location [EB/OL]. [2019－05－20]. https://doi.org/10.1007/s10796-018-9889-9Is.

[87] SHAH T A, GUL S, GAUR R C. Authors self-citation behavior in the field of library and information science[J]. Aslib journal of information management, 2015,67(4):458－468.

[88] ROUSSEAU R. Temporal differences in self-citation rates of scientific journals[J]. Scientometrics,1999,44(3):521－531.

[89] LIN W Y L, HUANG M H. The relationship between co-authorship, currency of references and author self-citations[J]. Scientometrics,2012,90(2):343－360.

[90] 魏瑞斌. 基于自引网络和内容分析的学者研究主题挖掘[J]. 情报学报,2015,34(6):635－645.

[91] HELLSTEN I, LAMBIOTTE R, SCHARNHORST A, et al. Self-citations, co-authorships and keywords: a new approach to scientists' field mobility? [J]. Scientometrics, 2007,72(3):469－486.

[92] LEE J Y. Exploring a researcher's personal research history through self-citation network and citation identity[J]. Journal of the Korean Society for Information Management,2012,29(1):157－174.

[93] 魏瑞斌. 基于自引网络和主路径分析的论文主题创新实证研究[J]. 图书情报工作,2018,62(3):64－70.

[94] EL-AICHOUCHI A, GORRY P. Delayed recognition of Judah Folkman's hypothesis on tumor angiogenesis: when a prince awakens a sleeping beauty by self-citation[J]. Scientometrics, 2018, 116(1):385－399.

[95] SEEBER M, CATTANEO M, MEOLI M, et al. Self-citations as strategic response to the use of metrics for career decisions[J]. Research policy, 2019,48(2):478－491.

[96] MCCUEN R H. Advancing scientific knowledge: ethical issues in the journal publication process[J]. Publicaions, 2018,6(1):1－6.

[97] SCARPA F, BIANCO V, TAGLIAFICO L A. The impact of the national assessment exercises on self-citation rate and publication venue: an empirical investigation on the engineering academic sector in Italy[J]. Scientometrics,2018,117(2):997－1022.

[98] KHALED M. Aberration of the citation[J]. Accountability in research, 2016,23 (4): 230－244.

[99] RACZ A, MARKOVIC S. "Worth(less) papers"-are journal impact factor and number of citations suitable indicators to evaluate quality of scientists? [J]. Nova prisutnost, 2018,16(2):369－389.

[100] 康存辉,操菊华. 期刊评价之自引辩解[J]. 编辑之友,2014(10):21－23,26.

Review of Self-citation Research: The Questioning, Verification and Innovation in Scientific Evaluation and Scientific Communication

Wen Fangfang
School of Management, Henan University of Science and Technology, Luoyang 471023

Abstract: [**Purpose/significance**] Self-citation is prevalent in scientific communication. It has been highly disputed during the recent decades. The review of self-citation research can enhance people's understanding and clarify their bias and misunderstanding on self-citation, as well as inspire follow-up researches and reflections. [**Method/process**] Through the systematic investigation of domestic and foreign literatures, this paper objectively describes the research history and current situation, displays its evolutionary trajectory, summarizes main results and ideas, illustrates the problems existing in current research, and predicts the focus and direction of future research. [**Result/conclusion**] Self-citation research has experienced long-term dispute and repeated verification without a unified conclusion. The dispute stems from the difference in research perspectives. Self-citation research needs new breakthroughs, with the focus turning from simple statistics and bibliometrics to the profound analysis of the laws and mechanisms behind data. In addition, the functions of self-citation in examining academic inheritance and knowledge diffusion will be further explored.

Keywords: self-citation scientific communication scientific evaluation citation-based indicators